

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
8 April 2004 (08.04.2004)

PCT

(10) International Publication Number  
WO 2004/029235 A2

- (51) International Patent Classification<sup>7</sup>: C12N 9/02, 15/52, 5/10
- (21) International Application Number: PCT/EP2003/010498
- (22) International Filing Date: 22 September 2003 (22.09.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 02021650.3 27 September 2002 (27.09.2002) EP
- (71) Applicant (for all designated States except US): DSM IP ASSETS B.V. [NL/NL]; Het Overloon 1, NL-6411 TE Heerlen (NL).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): HOSHINO, Tatsuo [JP/JP]; 2-18-14 Fueta, Kamakura-shi, Kanagawa-ken 248-0027 (JP). MIYAZAKI, Taro [JP/JP]; Kameino 2-14-8, Fujisawa-shi, Kanagawa-ken 252-0813 (JP). SUGISAWA, Teruhide [JP/CH]; Inzlingerstrasse 80, CH-4125 Riehen (CH).
- (74) Agent: MUELLER, Ingrid; Roche Vitamins Ltd., Patent Department (VMD), Wurmisweg 576, CH-4303 Kaiseraugst (CH).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
- without international search report and to be republished upon receipt of that report
  - with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 2004/029235 A2

(54) Title: ALDEHYDE DEHYDROGENASE GENE

(57) Abstract: The present invention relates to a novel DNA which encodes aldehyde dehydrogenase (SNDH), an expression vector containing the DNA and recombinant organisms containing said DNA. Furthermore, the present invention concerns a process for producing recombinant aldehyde dehydrogenase protein and a process for producing L-ascorbic acid (vitamin C) and/or 2-keto-L-gulonic acid (2-KGA) from L-sorbose by using the recombinant aldehyde dehydrogenase protein or recombinant organisms containing the expression vector. Also provided is a process for the production of 2KGA with a microorganism in which the gene encoding said aldehyde dehydrogenase is disrupted.

Aldehyde dehydrogenase gene

The present invention relates to a novel DNA which encodes aldehyde dehydrogenase (SNDH) derived from *Gluconobacter oxydans* DSM 4025, an expression vector containing the DNA and a recombinant microorganism containing the expression vector. Furthermore, the present invention concerns a process for producing  
5 recombinant aldehyde dehydrogenase protein and a process for producing L-ascorbic acid (vitamin C) and/or 2-keto-L-gulonic acid (2-KGA) from L-sorbose by using the recombinant aldehyde dehydrogenase protein or the recombinant microorganism containing said expression vector.

Vitamin C is one of indispensable nutrient factors for human beings and has been  
10 commercially synthesized by the Reichstein process for about 60 years. Synthetic vitamin C is also used in animal feed even though farm animals can synthesize it in their own body. Although the Reichstein process has many advantageous points for industrial vitamin C production, it still has undesirable problems such as high energy consumption and usage of considerable quantities of organic and inorganic solvents. Therefore, over  
15 the past decades, many approaches to manufacture vitamin C using enzymatic conversions, which would be more economical as well as ecological, have been investigated.

The present invention is directed to an isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide being at least 95% identical to  
20 the nucleotide sequence of SEQ ID NO: 1.

As used herein, "SNDH" stands for aldehyde dehydrogenase.

As used herein, "nucleic acid molecule" includes both DNA and RNA and, unless otherwise specified, includes both double-stranded, single-stranded nucleic acid, and nucleosides thereof. Also included are hybrids such as DNA-RNA hybrids, DNA-RNA-  
25 protein hybrids, RNA-protein hybrids, and DNA-protein hybrids.

As used herein, "mutation" refers to a single base pair change, insertion or deletion in the nucleotide sequence of interest.

- 2 -

As used herein, "mutagenesis" refers to a process whereby a mutation is generated in the DNA. With "random" mutagenesis, the exact site of mutation is not predictable, occurring anywhere in the chromosome of the microorganism, and the mutation is brought about as a result of physical damage caused by agents such as radiation or  
5 chemical treatment.

As used herein, "promoter" means a DNA sequence generally described as the 5' region of a gene, located proximal to the start codon. The transcription of the adjacent gene(s) is initiated at the promoter region. If a promoter is an inducible promoter, then the rate of transcription increases in response to an inducing agent. In contrast, the rate  
10 of transcription is not regulated by an inducing agent if the promoter is a constitutive promoter.

As used herein, "percent identical" refers to the percent of the nucleotides or amino acids of the subject nucleotide or amino acid sequence that have been matched to identical nucleotides or amino acids in the compared nucleotide or amino acid sequence  
15 by a sequence analysis program as exemplified below.

The present invention includes an isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide being at least 95% identical to the polynucleotide selected from the group consisting of (a) nucleotides 258-2084 of SEQ ID NO: 1, (b) nucleotides 351-2084 of SEQ ID NO: 1, (c) nucleotides 258-1955 of SEQ  
20 ID NO: 1, and (d) nucleotides 351-1955 of SEQ ID NO: 1.

It is another aspect of the present invention to provide an isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide selected from the group consisting of (a) a polynucleotide encoding the polypeptide having the amino acid sequence of SEQ ID NO: 2, (b) a polynucleotide encoding the polypeptide  
25 consisting of amino acids 32-609 of SEQ ID NO: 2, (c) a polynucleotide encoding the polypeptide consisting of amino acids 1-566 of SEQ ID NO: 2, and (d) a polynucleotide encoding the polypeptide consisting of amino acids 32-566 of SEQ ID NO: 2. Also included are proteins having SNDH activity and which are derived from a protein mentioned above by substitution, deletion, insertion or addition of one or more amino  
30 acid(s) in the amino acid sequences mentioned above.

Functional derivatives as another aspect of the present invention are defined on the basis of the amino acid sequences of the present invention by addition, insertion, deletion and/or substitution of one or more amino acid residues of such sequences wherein such

derivatives still have the SNDH activity measured by an assay known in the art or specifically described herein. Such functional derivatives can be made either by chemical peptide synthesis known in the art or by recombinant techniques on the basis of the DNA sequences as disclosed herein by methods known in the state of the art. Amino acid  
5 exchanges in proteins and peptides which do not generally alter the activity of such molecules are known in the state of the art.

In particular embodiments of the present invention, conservative substitutions of interest occur as follows: as example substitutions, Ala to Val/Leu/Ile, Arg to Lys/Gln/Asn, Asn to Gln/His/Lys/Arg, Asp to Glu, Cys to Ser, Gln to Asn, Glu to Asp, Gly  
10 to Pro/Ala, His to Asn/Gln/Lys/Arg, Ile to Leu/Val/Met/Ala/Phe/norLeu, Lys to Arg/Gln/Asn, Met to Leu/Phe/Ile, Phe to Leu/Val/Ile/Ala/Tyr, Pro to Ala, Ser to Thr, Thr to Ser, Trp to Tyr/Phe, Tyr to Trp/Phe/Thr/Ser, and Val to Ile/Leu/Met/Phe/Ala/norLeu are reasonable. As preferred examples, Ala to Val, Arg to Lys, Asn to Gln, Asp to Glu, Cys to Ser, Gln to Asn, Glu to Asp, Gly to Ala, His to Arg, Ile to Leu, Leu to Ile, Lys to Arg,  
15 Met to Leu, Phe to Leu, Pro to Ala, Ser to Thr, Thr to Ser, Trp to Tyr, Tyr to Phe, and Val to Leu are reasonable. If such substitutions result in a change in biological activity, then more substantial changes, denominated exemplary substitutions described above, are introduced and the products screened.

Furthermore the present invention is directed to polynucleotides encoding  
20 polypeptides having the SNDH activity as disclosed in the sequence listing as SEQ ID NO: 2 as well as the complementary strands, or those which include these sequences, DNA sequences or fragments thereof, and DNA sequences, which hybridize under standard conditions with such sequences but which encode for polypeptides having exactly the same amino acid sequence.

25 Thus, the present invention provides an isolated nucleic acid molecule encoding a polypeptide having aldehyde dehydrogenase activity, wherein the complement of said nucleic acid molecule hybridizes under standard conditions with a nucleic acid molecule as described above. It is an aspect of the invention to provide an isolated nucleic acid molecule encoding a polypeptide having aldehyde dehydrogenase activity, wherein said  
30 nucleic acid molecule hybridizes under standard conditions to the complementary strand of a nucleic acid molecule encoding (i) aldehyde dehydrogenase which comprises a polynucleotide being at least 95% identical to the nucleotide sequence of SEQ ID NO: 1; (ii) aldehyde dehydrogenase which comprises a polynucleotide being at least 95% identical to the polynucleotide selected from the group consisting of (a) nucleotides 258-  
35 2084 of SEQ ID NO: 1, (b) nucleotides 351-2084 of SEQ ID NO: 1, (c) nucleotides 258-

1955 of SEQ ID NO: 1, and (d) nucleotides 351-1955 of SEQ ID NO: 1; and (iii) aldehyde dehydrogenase which comprises a polynucleotide selected from the group consisting of (a) a polynucleotide encoding the polypeptide having the amino acid sequence of SEQ ID NO: 2, (b) a polynucleotide encoding the polypeptide consisting of amino acids 32-609  
5 of SEQ ID NO: 2, (c) a polynucleotide encoding the polypeptide consisting of amino acids 1-566 of SEQ ID NO: 2, and (d) a polynucleotide encoding the polypeptide consisting of amino acids 32-566 of SEQ ID NO: 2.

"Standard conditions" for hybridization mean in this context the conditions which are generally used by a person skilled in the art to detect specific hybridization signals, or  
10 preferably so called stringent hybridization conditions used by a person skilled in the art.

Thus, as used herein, the term "stringent hybridization conditions" means that hybridization will occur if there is 95% and preferably at least 97% identity between the sequences. Stringent hybridization conditions are, e.g., conditions under over night incubation at 42°C using a digoxigenin (DIG)-labeled DNA probe (constructed by using  
15 a DIG labeling system; Roche Diagnostics GmbH, 68298 Mannheim, Germany) in a solution comprising 50% formamide, 5 × SSC (150 mM NaCl, 15 mM trisodium citrate), 0.2% sodium dodecyl sulfate, 0.1% N-lauroylsarcosine, and 2% blocking reagent (Roche Diagnostics GmbH), followed by washing the filters in 0.1× SSC at about 60°C.

This invention is also directed to a recombinant vector, i.e., an expression vector,  
20 comprising such a nucleic acid molecule as mentioned above. An expression vector of the present invention is one which functions in a suitable host cell. Preferred vectors for the expression of the nucleic acid molecules of the present invention are vectors or derivatives thereof which are selected from the group consisting of pQE, pUC, pBluescript II, pACYC177, pACYC184, pVK100, and RSF1010.

25 A suitable host cell for expression of the nucleotide sequences of the present invention is a recombinant microorganism selected from the group consisting of bacteria, yeast, and plant cells. Preferably, the microorganism is selected from the group consisting of *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* and *Escherichia*. An example of such a preferred microorganism is *E. coli*. A more preferred  
30 host cell belongs to *Gluconobacter oxydans*, most preferably *G. oxydans* DSM 4025 (FERM BP-3812), which had been deposited on March 17, 1987 under the conditions of the Budapest Treaty at the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany.

The microorganism "*Gluconobacter oxydans*" also includes synonyms or basonyms of such species having the same physico-chemical properties, as defined by the International Code of Nomenclature of Prokaryotes.

Thus, the present invention is directed to a recombinant microorganism which is transformed with the expression vector as described above or which comprises a nucleic acid molecule as described above integrated into its chromosomal DNA.

A wide variety of host/vector combinations may be used for cloning the double stranded nucleotide sequences of the present invention. As *E. coli* is a preferred host cell, any vectors usually used in *E. coli* are useful for the present invention. Such vectors include, but are not limited to, pQE vectors which can express His-tagged recombinant proteins (QIAGEN K.K., Tokyo, Japan), pBR322 or its derivatives including pUC18 and pBluescript II (Stratagene Cloning Systems, California, USA), pACYC177 and pACYC184 and their derivatives, and a vector derived from a broad host range plasmid such as RK2 and RSF1010. Thus, the expression vector used in the present invention is derived from pQE-plasmids, pUC-plasmids, pBluescript II, pACYC177, pACYC184, and their derivative plasmids, and a broad host range plasmid such as pVK100 and RSF1010.

As used herein, "expression vector" means a cloning vector which is capable of enhancing the expression of a gene that has been cloned into it, after transformation into a suitable host. The cloned gene is usually placed under the control of (i.e., operably linked to) certain control sequences such as promoter sequences. Promoter sequences may be either constitutive or inducible.

As used herein, "cloning vector" means a plasmid or phage DNA or other DNA sequence which is able to replicate autonomously in a host cell, and which is characterized by a single or a small number of restriction endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without loss of an essential biological function of the vector, and into which a DNA fragment may be introduced in order to bring about its replication and cloning. The cloning vector may further contain a marker suitable for use in the identification of cells transformed with the cloning vector. Such markers provide, e.g., tetracycline or ampicillin resistance.

As used herein, a "recombinant vector" includes any cloning or expression vector which contains the desired cloned gene(s).

As used herein, "expression" refers to the process by which a polypeptide is produced from a structural gene. The process involves transcription of the gene into mRNA and the translation of such mRNA into polypeptide(s).

As used herein, "recombinant microorganism" includes a recombinant host which  
5 may be any prokaryotic or eukaryotic cell that contains the desired cloned gene(s) on an expression or cloning vector. This term also includes those prokaryotic or eukaryotic cells that have been genetically engineered to contain the desired gene(s) in the chromosome or genome of said microorganism.

As used herein, "host" includes any prokaryotic or eukaryotic cell that is the  
10 recipient of a replicable expression vector or cloning vector. A "host", as the term is used herein, also includes prokaryotic or eukaryotic cells that can be genetically engineered by well known techniques to contain desired gene(s) on its chromosome or genome. Examples of such hosts are known to the skilled artisan.

To construct a recombinant microorganism carrying recombinant DNA, e.g., a  
15 recombinant vector, various gene transfer methods including, but not limited to, transformation, transduction, conjugal mating or electroporation can be used. These methods are well-known in the field of molecular biology. Conventional transformation systems can be used for *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*. A transduction system can also be used for *E. coli*. Conjugal mating  
20 systems can be widely used in Gram-positive and Gram-negative bacteria including *E. coli*, *P. putida*, and *Gluconobacter*. An example of conjugal mating is disclosed in WO 89/06,688. The conjugation can occur in liquid medium or on a solid surface. Examples for a suitable recipient for SNDH production include microorganisms of *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*. To the recipient for  
25 conjugal mating, a selective marker may be added, e.g., resistance against nalidixic acid or rifampicin. Natural resistance can also be used, e.g., resistance against polymyxin B is useful for many *Gluconobacters*.

Preferred vectors useful for the present invention are broad-host-range vectors such as a cosmid vector like pVK100 and its derivatives and RSF1010. Copy number and  
30 stability of the vector should be carefully considered for stable and efficient expression of the cloned nucleic acid molecules and also for efficient cultivation of the host cell carrying said cloned molecules. Nucleic acid molecules containing transposable elements such as Tn5 can also be used to introduce the desired DNAs into the preferred host, especially on a chromosome. Nucleic acid molecules containing any DNAs isolated

- 7 -

from the preferred host together with the nucleotide sequences of the present invention are also useful to introduce the nucleotide sequences of the present invention into the preferred host cell, especially on a chromosome. Such nucleic acid molecules can be transferred to the preferred host by applying any of a conventional method, e.g.,  
5 transformation, transduction, conjugal mating or electroporation, which are well known in the art, considering the nature of the host cell and the nucleic acid molecule.

The nucleotide sequences including the SNDH gene provided in this invention are ligated into a suitable vector containing a regulatory region such as a promoter, a ribosomal binding site, and a transcriptional terminator operable in the host cell  
10 described above with a method well-known in the art to produce a suitable expression vector.

To express the desired gene/nucleotide sequence isolated from *G. oxydans* DSM 4025 efficiently, various promoters can be used; e.g., the original promoter of the gene, promoters of antibiotic resistance genes such as kanamycin resistant gene of Tn5,  
15 ampicillin resistant gene of pBR322, and beta-galactosidase of *E. coli* (lac), trp-, tac-, trc-promoter, promoters of lambda phage and any promoters which are functional in a host cell. For this purpose, the host cell can be selected from the group consisting of bacteria, yeast, and plant cells. Preferably, the host cell belongs to the genera *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*.

20 For expression, other regulatory elements, such as a Shine-Dalgarno (SD) sequence (e.g., AGGAGG, including natural and synthetic sequences operable in the host cell) and a transcriptional terminator (inverted repeat structure including any natural and synthetic sequence operable in the host cell) which are operable in the host cell (into which the coding sequence will be introduced to provide a recombinant cell of this invention)  
25 can be used with the above described promoters.

For the expression of polypeptides which are located in the periplasmic space, like the SNDH protein of the present invention, a signal peptide, which contains usually 15 to 50 amino acid residues and is totally hydrophobic, is preferably associated. A DNA encoding a signal peptide can be selected from any natural and synthetic sequence  
30 operable in the desired host cell. A putative signal peptide containing amino acid residues 1-31 of SEQ ID NO: 2 was also found in the protein expressed by the SNDH gene of the present invention (SEQ ID NO: 4).

- 8 -

Unless otherwise mentioned, all amino acid sequences determined by sequencing the purified SNDH protein herein were determined using an automated amino acid sequencer (such as model 470A, Perkin-Elmer Applied Biosystems).

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer (such as the model ALF express II, Amersham Pharmacia Biotech), and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of the DNA sequence determined as above. Therefore, as it is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

The invention provides an isolated nucleic acid molecule encoding the enzyme (SNDH). Methods and techniques designed for the manipulation of isolated nucleic acid molecules are well known in the art. Methods for the isolation, purification, and cloning of nucleic acid molecules, as well as methods and techniques describing the use of eukaryotic and prokaryotic host cells and nucleic acid and protein expression therein, are known to the skilled person.

Briefly, the SNDH gene, the DNA molecule containing said gene, the recombinant expression vector and the recombinant microorganism used in the present invention can be obtained by the following steps:

- (1) isolating chromosomal DNA from *G. oxydans* DSM 4025 and constructing a gene library with the chromosomal DNA in an appropriate host cell, e.g., *E. coli*;
- (2) cloning the SNDH gene from the chromosomal DNA by colony-, plaque-, or Southern-hybridization, PCR (polymerase chain reaction) cloning, western-blot analysis or other techniques known in the art;

- 9 -

- (3) determining the nucleotide sequence of the SNDH gene obtained as above by conventional methods to select a DNA molecule containing said SNDH gene and constructing a recombinant expression vector on which the SNDH gene can be expressed efficiently;
- 5 (4) constructing recombinant microorganisms carrying the SNDH gene by an appropriate method for introducing DNA into a host cell, e.g., transformation, transduction, conjugal transfer and/or electroporation, which host cell thereby becomes a recombinant microorganism of this invention.

The materials and techniques used in the above aspect of the present invention are  
10 exemplified in detail as follows:

A total chromosomal DNA can be purified by a procedure well known in the art. The desired gene can be cloned in either plasmid or phage vectors from a total chromosomal DNA typically by either of the following illustrative methods:

- (i) The partial amino acid sequences are determined from the purified proteins or  
15 peptide fragments thereof. Such whole protein or peptide fragments can be prepared by the isolation of such a whole protein or by peptidase-treatment from the gel after SDS-polyacrylamide gel electrophoresis. Thus obtained protein or fragments thereof are applied to protein sequencer such as Applied Biosystems automatic gas-phase sequencer 470A. The amino acid sequences can be utilized to design and prepare oligonucleotide  
20 probes and/or primers with DNA synthesizer such as Applied Biosystems automatic DNA sequencer 381A. The probes can be used for isolating clones carrying the target gene from a gene library of the strain carrying the target gene by means of Southern-, colony- or plaque-hybridization.
- (ii) Alternatively, for the purpose of selecting clones expressing a target protein from the  
25 gene library, immunological methods with antibodies prepared against the target protein can be applied.
- (iii) The DNA fragment of the target gene can be amplified from the total chromosomal DNA by PCR method with a set of primers, i.e., two oligonucleotides synthesized according to the amino acid sequences determined as above. Then a clone carrying the  
30 target-whole gene can be isolated from the gene library constructed, e.g., in *E. coli*, by Southern-, colony-, or plaque-hybridization with the PCR product obtained above as probe.

DNA sequences which can be made by PCR by using primers designed on the basis of the DNA sequences disclosed herein by methods known in the art are also an object of  
35 the present invention.

- 10 -

Above mentioned antibodies can be prepared with the purified SNDH proteins, the purified recombinant SNDH proteins such as His-tagged SNDH expressed in *E. coli*, or its peptide fragment as an antigen.

Once a clone carrying the desired gene is obtained, the nucleotide sequence of the  
5 target gene can be determined by a well known method such as dideoxy chain  
termination method with M13 phage.

#### Brief description of the Figures:

Figure 1 illustrates the gene encoding a protein having aldehyde dehydrogenase activity of the present invention. The restriction map of the SNDH gene and ORF-A  
10 gene is given, wherein ORF means open reading frame and Signal seq. means the putative  
signal peptide sequence of the SNDH gene.

Figure 2 illustrates the restriction map of the SNDH and ORF-A genes cloned into cosmid pVSN5 as well as cloning of insert DNA of different sizes into pUC plasmids pUCSNP4, pUCSNP9, pUCSN19, and pUCSN5. In the physical map of pVSN5, the  
15 arrow filled in gray shows the SNDH gene.

Figure 3 shows the cloning strategy of the 8.0 kb *Pst*I fragment including the intact SNDH gene from pVSN5 into pUC18 vectors to result in pUCSNP4 and pUCSNP9. Note that pUCSNP4 and pUCSNP9 are identical except for the direction of the insert.

Figure 4 shows schematically the construction of the GOMTR1SN::Km (SNDH-  
20 disruptant) by using a suicide vector plasmid having the disrupted SNDH gene with  
kanamycin cassette (Km). Homologous recombination between the vector plasmid and  
the chromosomal DNA of GOMTR1 as the parent strain at the corresponding region  
occurs to obtain the disruptant strains. "G.O." means *Gluconobacter oxydans*.

Figure 5 represents the physical map of the insert DNA of pVSN117, pVSN106, and  
25 pVSN114. Plasmid pVSN117 has the insert DNA encoding the C-terminal deleted  
SNDH gene (nucleotides 258-1955 of SEQ ID NO: 1, i.e., amino acids 1-566 of SEQ ID  
NO: 2), which expresses only a 55 kDa protein. Plasmids pVSN106 and pVSN114 have  
the insert DNA encoding the intact SNDH gene.

The gene of the present invention encodes an SNDH enzyme of 578 amino acid  
30 residues (SEQ ID NO: 5 consisting of amino acids 32-609 of SEQ ID NO: 2) together  
with a putative signal peptide of 31 amino acid residues (SEQ ID NO: 4 consisting of

amino acids 1-31 of SEQ ID NO: 2) as depicted in Figure 2. In terms of nucleotide sequences, the coding region of the SNDH gene encompasses nucleotides 258-2087 of SEQ ID NO: 1 and includes the coding sequences for a putative signal peptide (nucleotides 258-350 of SEQ ID NO: 1) and the stop codon (nucleotides 2085-2087 of  
5 SEQ ID NO: 1). Thus, the nucleotide sequence without the stop codon is the nucleotide sequence from position 258-2084 of SEQ ID NO: 1, and additionally without the signal sequence the nucleotide sequence encompasses nucleotides 351-2084 of SEQ ID NO: 1.

The nucleic acid molecules as disclosed in the present invention can be used in a recombinant microorganism for the production of 2-KGA and/or vitamin C. The  
10 recombinant microorganism, which is selected from *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* and *Escherichia*, may be cultured in an aqueous medium supplemented with appropriate nutrients under aerobic conditions. The cultivation may be conducted at a pH of 4.0 to 9.0, preferably 6.0 to 8.0. The cultivation period varies depending on the pH, temperature and nutrient medium to be used, and is  
15 preferably about 1 to 5 days. The preferred temperature range for carrying out the cultivation is from about 13°C to about 36°C, preferably from about 18°C to about 33°C. It is usually required that the culture medium contains such nutrients as assimilable carbon sources, e.g., glycerol, D-mannitol, D-sorbitol, erythritol, ribitol, xylitol, arabitol, inositol, dulcitol, D-ribose, D-fructose, D-glucose, and sucrose, preferably D-sorbitol, D-  
20 mannitol, and glycerol; and digestible nitrogen sources such as organic substances, e.g., peptone, yeast extract, baker's yeast, urea, amino acids, and corn steep liquor. Various inorganic substances may also be used as nitrogen sources, e.g., nitrates and ammonium salts. Furthermore, the culture medium usually contains inorganic salts, e.g., magnesium sulfate, potassium phosphate, and calcium carbonate. The cultivation is carried out in  
25 appropriate equipment such as jar fermentors, flasks, or tubes. The recombinant microorganism is either transformed with an expression vector containing the nucleic acid molecules as above or comprises the nucleic acid molecule of the present invention integrated into its chromosomal DNA.

An appropriate sugar compound used as substrate for the production of 2-KGA  
30 and/or vitamin C is L-sorbose. The metabolic pathway for 2-KGA and vitamin C goes from D-sorbitol via L-sorbose to L-sorbose, which is then converted to 2-KGA and/or vitamin C. Thus, the direct substrate for both products is L-sorbose.

Thus, it is an aspect of the present invention to provide a process for the production of 2-KGA and/or vitamin C from L-sorbose comprising (a) propagating or  
35 cultivating the recombinant microorganism, which is either transformed with an

the present invention or which comprises the nucleic acid molecules of the present invention integrated into its chromosomal DNA, in an appropriate culture medium and (b) recovering and separating 2-KGA and/or vitamin C from said culture medium.

It is one embodiment to provide a process for the production of vitamin C and/or 2-KGA from L-sorbose comprising (a) propagating a recombinant organism in an appropriate culture medium, wherein the nucleic acid molecule as of the present invention is heterologously introduced to said recombinant organism, and (b) recovering and separating vitamin C and/or 2-KGA from said culture medium.

The present invention provides recombinant SNDH. One can increase the production yield of the SNDH enzyme by introducing the SNDH gene provided by the present invention into a host cell including *G. oxydans* DSM 4025. One can also produce more efficiently the SNDH proteins in a host cell selected from a group consisting of *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* and *Escherichia* by using the SNDH gene of the present invention. The microorganism may be cultured as described above.

An embodiment for the isolation and purification of the recombinant SNDH from the microorganism after the cultivation is briefly described hereinafter: cells are harvested from the liquid culture broth by centrifugation or filtration. The harvested cells are washed with water, physiological saline or a buffer solution having an appropriate pH. The washed cells are suspended in the buffer solution and disrupted by means of a homogenizer, sonicator or French press, or by treatment with lysozyme to give a solution of disrupted cells. The recombinant SNDH is isolated and purified from the cell-free extract of disrupted cells, preferably from the cytosol fraction of the microorganism. The recombinant SNDH can be immobilized on a solid carrier for solid phase enzyme reaction.

The invention is further directed to a process for the production of 2-KGA from L-sorbose comprising (a) cultivating a microorganism belonging to *Gluconobacter oxydans* DSM 4025 in an appropriate culture medium, wherein the gene encoding aldehyde dehydrogenase represented by SEQ ID NO: 2 is disrupted in said microorganism, and (b) recovering and separating 2-KGA from said culture medium. The disruption may take place anywhere in the gene, resulting in a non-functioning of the encoded enzyme.

Thus, a process is provided for the production of 2-KGA via L-sorbose from an appropriate sugar compound comprising (a) propagating a microorganism belonging to *Gluconobacter oxydans* DSM 4025 in an appropriate culture medium, wherein the gene encoding aldehyde dehydrogenase is disrupted in said microorganism, said aldehyde dehydrogenase being encoded by (i) a polynucleotide being at least 95% identical to the nucleotide sequence of SEQ ID NO: 1; (ii) a polynucleotide being at least 95% identical to the polynucleotide selected from the group consisting of (a) nucleotides 258-2084 of SEQ ID NO: 1, (b) nucleotides 351-2084 of SEQ ID NO: 1, (c) nucleotides 258-1955 of SEQ ID NO: 1, and (d) nucleotides 351-1955 of SEQ ID NO: 1; and (iii) a polynucleotide selected from the group consisting of (a) a polynucleotide encoding the polypeptide having the amino acid sequence of SEQ ID NO: 2, (b) a polynucleotide encoding the polypeptide consisting of amino acids 32-609 of SEQ ID NO: 2, (c) a polynucleotide encoding the polypeptide consisting of amino acids 1-566 of SEQ ID NO: 2, and (d) a polynucleotide encoding the polypeptide consisting of amino acids 32-566 of SEQ ID NO: 2. The resulting 2-KGA is further recovered and isolated from said culture medium.

In one embodiment, the invention provides a process for the disruption of the SNDH gene by classical mutagenesis with agents such as UV-irradiation or chemical treatment by any mutation reagents, e.g., *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (NTG), ICR170 or acridine orange, *in vivo* as well as *in vitro*.

In another embodiment, the invention provides a process for the disruption of the SNDH gene by DNA recombination techniques such as transposon insertion or site directed mutagenesis by PCR, *in vivo* as well as *in vitro*.

In another embodiment, the invention provides a process for producing 2-KGA using the disruptants described above by fermentation from an appropriate substrate, i.e., a sugar compound, which is selected from the group consisting of L-sorbose, D-glucose, D-sorbitol, and L-sorbose. The process takes place in appropriate equipment such as jar fermentors, flasks, or tubes. Furthermore, the invention provides a process for producing 2-KGA using a cell free extract of the disruptants described above by incubation from an appropriate substrate, e.g., L-sorbose, D-glucose, L-sorbose, and D-sorbitol, in appropriate equipment such as a bioreactor.

The present invention provides recombinant SNDH. Furthermore, it is directed to a process for the production of aldehyde dehydrogenase comprising (a) cultivating a recombinant microorganism comprising a nucleic acid molecule encoding aldehyde dehydrogenase which comprises (i) a polynucleotide being at least 95% identical to the

nucleotide sequence of SEQ ID NO: 1; (ii) a polynucleotide being at least 95% identical to the polynucleotide selected from the group consisting of (a) nucleotides 258-2084 of SEQ ID NO: 1, (b) nucleotides 351-2084 of SEQ ID NO: 1, (c) nucleotides 258-1955 of SEQ ID NO: 1, and (d) nucleotides 351-1955 of SEQ ID NO: 1; and (iii) a polynucleotide  
5 selected from the group consisting of (a) a polynucleotide encoding the polypeptide having the amino acid sequence of SEQ ID NO: 2, (b) a polynucleotide encoding the polypeptide consisting of amino acids 32-609 of SEQ ID NO: 2, (c) a polynucleotide encoding the polypeptide consisting of amino acids 1-566 of SEQ ID NO: 2, and (d) a polynucleotide encoding the polypeptide consisting of amino acids 32-566 of SEQ ID  
10 NO: 2; wherein said microorganism is cultivated in an appropriate culture medium and (b) wherein said aldehyde dehydrogenase is recovered and separated from said culture medium.

**Example 1: Amino acid sequencing from the N-terminus of SNDH**

15 The partial amino acid sequence of the N-terminal 75 kDa subunit of the SNDH protein was determined. About 10 µg of the SDS-treated purified SNDH consisting of 75 kDa subunits was subjected to SDS-PAGE, and the protein band was electroblotted onto a PVDF membrane. The protein blotted on the membrane was soaked in a digestion buffer (100 mM potassium phosphate buffer, 5 mM dithiothreitol, 10 mM EDTA, pH  
20 8.0) and incubated with 5.04 µg of pyroglutamate aminopeptidase (SIGMA, USA) at 30°C for 24 hours. After incubation, the membrane was washed with deionized water and subjected to N-terminal amino acid sequencing using an automated amino acid sequencer (ABI model 490, Perkin Elmer Corp., Conn., USA). As a result, 14 residues of the N-terminal amino acid sequence were obtained as illustrated in SEQ ID NO: 3.

25

**Example 2: Cloning of partial SNDH gene by PCR**

Amplification of the partial SNDH gene fragment was carried out by PCR with chromosomal DNA of *G. oxydans* DSM 4025 (FERM BP-3812) and degenerated oligonucleotide DNA primers, P11 (SEQ ID NO: 6) and P12 (SEQ ID NO: 7). Both  
30 primers were degenerated DNA mixtures having bias for *Gluconobacter* codon usage. The PCR was performed with thermostable taq polymerase (TAKARA Ex Taq™, Takara Shuzo Co., Ltd., Seta 3-4-1, Otsu, Shiga, 520-2193, Japan), using a thermal cycler (Gene

Amp PCR System 2400-R, PE Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA). The reaction mixture (25 µl) consisted of 200 µM of dNTPs, 50 pmol of each primer (24 ~ 48 degeneracy), 5 ng of the chromosomal DNA, and 1.25 units of the DNA polymerase in the buffer provided from the supplier. The reaction was carried out  
5 with 5 cycles of denaturation step at 94°C for 30 sec, annealing step at 37°C for 30 sec, synthesis step at 70°C for 1 min followed by 25 cycles of denaturation step at 94°C for 30 sec, annealing step at 50°C for 30 sec, synthesis step at 70°C for 1 min. As a result, a 41 bp DNA fragment was specifically amplified and cloned into vector pCR 2.1-TOPO (Invitrogen, 1600 Faraday Avenue Carlsbad, California 92008, USA) to obtain a  
10 recombinant plasmid pMTSN2. The nucleotide sequence of the cloned 41 bp DNA fragment which encodes an N-terminal partial amino acid sequence of the mature SNDH protein was confirmed by dideoxy-chain termination method (F. Sanger et al, Proc. Natl. Acad. Sci. USA, 74, 5463-5467, 1977).

### 15 Example 3: Complete cloning of the SNDH gene

#### (1) Construction of a gene library of *G. oxydans* DSM 4025

The chromosomal DNA of *G. oxydans* DSM 4025 was prepared from cells grown on M agar medium containing 5% D-mannitol, 1.75% corn steep liquor, 5% baker's yeast, 0.25% MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.5% CaCO<sub>3</sub> (Practical Grade), 0.5% urea, and 2.0% agar  
20 (pH 7.0), for 4 days at 27°C. The chromosomal DNA (4 µg) was partially digested with 4 units of *EcoR* I in 20 µl of reaction mixture. A portion (8 µl) of the sample containing partially-digested DNA fragments was separated by an electrophoresis using 1% agarose gel. Fragments ranging from 15 to 35 kb were cut out and chemically melted to recover the fragments using QIAEX II (QIAGEN Inc., 28159 Avenue Stanford, Valencia, CA  
25 91355, USA). The objective DNA fragments recovered were suspended in H<sub>2</sub>O. On the other hand, 2 µg of a cosmid vector pVK100 was completely digested with *EcoR* I and dephosphorylated of the 5'-ends by treating with bacterial alkaline phosphatase (*E. coli* C75) (Takara Shuzo). The treated pVK100 (220 ng) was ligated with the 15 - 35 kb *EcoR* I fragments (1 µg) using a ligation kit (Takara Shuzo) in 36 µl of reaction mixture. The  
30 ligated DNA which had been ethanol precipitated and resolved in appropriate volume of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) was used for *in vitro* packaging (Gigapack III Gold Packaging Extract, Stratagene, 11011 North Torrey Pines Road, La Jolla, CA 92037, USA) to infect *E. coli* VCS257, a host strain for the genomic library. As a

result, totally 400,000 - 670,000 clones containing about 25 kb-inserted DNA fragments were obtained.

(2) Complete cloning of the SNDH gene by colony hybridization

The probe that would be used for screening of the cosmid library described above  
5 to detect clones carrying the complete SNDH gene by colony hybridization method, was  
constructed. The 41 bp DNA fragment encoding the N-terminal amino acid sequence of  
SNDH was amplified and labeled by PCR-DIG labeling method (Roche Molecular  
Systems Inc., 1145 Atlantic Avenue, Alabama, CA94501, USA). PCR with plasmid  
pMTSN2 DNA as a template and oligonucleotide DNA primers, P13 (SEQ ID NO: 8)  
10 and P14 (SEQ ID NO: 9), was performed with thermostable taq polymerase (TAKARA  
Ex Taq™, Takara Shuzo Co., Ltd.), using a thermal cycler (Gene Amp PCR System 2400-  
R, PE Biosystems). The reaction was carried out with 25 cycles of denaturation step at  
94°C for 30 sec, annealing step at 55°C for 30 sec, synthesis step at 70°C for 1 min. Using  
the DIG-labeled probe, screening of the cosmid library (about 1,000 clones) by colony  
15 hybridization and chemiluminescent detection according to the method provided from  
the supplier (Roche Molecular Systems Inc., USA) was carried out. As a result, 3 positive  
clones were isolated and one of them was designated pVSN5, which carried about 25 kb  
insert DNA in pVK100 vector. From this, 25 kb DNA insert fragments of different sizes  
were further subcloned into pUC18 vectors (Figure 2): (1) a 3.2 kb *EcoR* I fragment  
20 comprising the upstream portion (the N-terminal part) of the SNDH gene resulting in  
pUCSN19, (2) a 7.2 kb *EcoR* I fragment comprising the downstream portion (the C-  
terminal part) of the SNDH gene resulting in pUCSN5, and (3) a 1.8 kb *Pst* I fragment  
comprising the intact or complete SNDH gene resulting in pUCSNP4 and pUCSNP9,  
respectively. Note that the inserts in pUSNP4 and pUSNP9 are the same but in opposite  
25 directions.

(3) Nucleotide sequencing of the SNDH gene

Plasmids pUCSN19, pUCSN5, and pUCSNP4 were used for nucleotide sequencing  
of a region including the SNDH gene or gene fragments. The determined nucleotide  
sequence (SEQ ID NO: 1 with 3,408 bp) revealed that the ORF of the SNDH gene (1,827  
30 bp; nucleotides 258-2084 of SEQ ID NO: 1) encoded a polypeptide of 609 amino acid  
residues (SEQ ID NO: 2). An additional ORF, ORF-A, was found downstream of the  
SNDH ORF as illustrated in Figure 1. ORF-A (1,101 bp; nucleotides 2214-3314 of SEQ  
ID NO: 1) encoded a polypeptide of 367 amino acids.

- 17 -

In the ORF of the SNDH gene, a signal peptide-like sequence (SEQ ID NO: 4 with 31 amino acids) is possibly included in the deduced amino acid sequence, which contains (i) many hydrophobic residues, (ii) positively-charged residues close to the N-terminus and (iii) Ala-Xaa-Ala motif for cleavage site of the signal sequence. The putative  
5 ribosome-binding site (Shine-Dalgarno, SD, sequence) for the SNDH gene was located at 6 bp upstream of the initiation codon (AGGAGA at nucleotide position 247-252 of SEQ ID NO: 1). Furthermore, a motif (Cys-Xaa-Xaa-Cys-His) defined as heme c binding site was found at position 530-534 of SEQ ID NO: 2. From the genetically analysis as shown above, the SNDH protein is thought to be one of quinoxemoproteins.

10 A homology search for the SNDH gene using the program of FASTA in GCG (Genetics Computer Group, Madison, WI, USA) revealed that Arg227, Asn228, Gln230, Gly246, and Asp251 of SEQ ID NO: 5 correspond to several highly conserved residues in the presumed active site of *A. calcoaceticus* GDH-B protein described by Oubrie et al. [J. Mol. Biol. 289:319-333 (1999)].

15

#### Example 4: Expression of the SNDH gene in *E. coli*

Plasmids pUCSNP4 and pUCSNP9 (Figure 3), containing the 8.0 kb *Pst* I-fragment with the intact, i.e., complete SNDH gene, were transformed into *E. coli* JM109 to confirm the expression and the activity of the SNDH proteins. The amount of vitamin C  
20 produced as the enzyme activity was measured at a wavelength of 264nm by a high performance liquid chromatography system (HPLC) which was composed with a UV detector (TOSOH UV8000; Tosoh, Japan), a dualpump (TOSOH CCPE; Tosoh), an integrator (Shimadzu C-R6A; Shimadzu, Japan) and a column (YMC-Pack Polyamine-II; 4.6 mm of inner diameter [i.d.] × 15 cm, YMC, U.S.A.).

25 The conversion activity of L-sorbose to vitamin C by using cytosol fraction of the recombinant *E. coli* was tested (Table 1). Cells were cultivated in LB medium optionally supplemented with 10 μM of PQQ and 1.0 mM of CaCl<sub>2</sub>. The cytosol fraction was prepared by ultracentrifugation (100,000 x g, 45 min) of the cell free extract in 50 mM potassium phosphate buffer (pH 7.0). The reaction mixture (100 μl) consisted of  
30 125 μg of cytosol fraction of the recombinant *E. coli*, 50 mM of L-sorbose, 1.0 mM of phenazine mesosulfate (PMS), with or without the addition of 1.0 μM of PQQ and 1.0 mM of CaCl<sub>2</sub> as cofactors, depending on the case. The enzyme reaction was carried out at 30°C for 30 min. The holo-SNDHs of the cells cultivated in LB medium containing 10

$\mu\text{M}$  of PQQ and 1.0 mM of  $\text{CaCl}_2$  produced vitamin C definitely under the defined reaction condition without the cofactors of PQQ and  $\text{CaCl}_2$ . By addition of the cofactors, the apo-SNDH expressed with pUCSNP4 and pUCSNP9 showed almost the same activity as those of the holo-enzyme.

5

Table 1: activity measurement of recombinant SNDH

Microorganism	PQQ and $\text{CaCl}_2$ in the medium	Specific activity (mU/mg Protein)	
		with PQQ and $\text{CaCl}_2$	without PQQ and $\text{CaCl}_2$
<i>E. coli</i> JM109/pUCSNP4	+	0.187	0.224
<i>E. coli</i> JM109/pUCSNP9	+	0.198	0.252
<i>E. coli</i> JM109/pUC18	+	0.000	0.000
<i>E. coli</i> JM109/pUCSNP4	-	0.155	0.000
<i>E. coli</i> JM109/pUCSNP9	-	0.176	0.000
<i>E. coli</i> JM109/pUC18	-	0.000	0.000
<i>G. oxydans</i> DSM 4025	-	0.026	0.026

One unit (U) of the enzyme was defined as the amount of enzyme, which produces 1.0 mg of vitamin C in the defined reaction.

#### 10 Example 5: Construction and cultivation of SNDH-gene disruptants of *G. oxydans* strains

Figure 4 shows the scheme for the construction of SNDH gene targeting vector, GOMTR1SN::Km (SNDH-disruptant). First, plasmid pSUPSN was constructed by a ligation of the 8.0 kb *Pst* I fragment containing the SNDH gene from plasmid pUCSNP4 with a suicide vector pSUP202 (for reference see Simon et al.; A broad host range mobilization system for in vitro genetic engineering: transposon mutagenesis in Gram negative bacteria, Biotechnology, 1, 784-791, 1983). Second, a kanamycin-resistant-gene cassette (Km cassette) was inserted into the *Eco*R I site of the SNDH gene cloned in plasmid pSUPSN to obtain plasmid pSUPSN::Km ( $\text{Km}^r\text{Tc}^r$ ). Then, plasmid pSUPSN::Km was introduced into GOMTR1, which was a rifampicin (Rif) resistant derived spontaneously from wildtype *G. oxydans* DSM 4025 strain, to result in SNDH-null mutants ( $\text{Km}^r\text{Rif}^r\text{Tc}^r$ ).

GOMTR1 was cultivated in a 200 ml flask containing 50 ml of T broth, which was composed of 30 g/l of Trypticase Soy Broth (BBL; Becton Dickinson and Company, Cockeysville, MD 21030, USA) and 3 g/l of yeast extract (Difco; Becton Dickinson Microbiology Systems, Becton Dickinson and Company, Sparks, MD 21152, USA) with  
5 100 µg/ml of rifampicin at 30°C overnight. *E. coli* HB101 (pRK2013) [D. H. Figurski, Proc. Natl. Acad. Sci. USA, 76, 1648-1652, 1979] and *E. coli* JM109 (pSUPSN::Km) were cultivated in test tubes containing 2 ml of LB medium with 50 µg/ml of kanamycin at 30°C overnight. Cultured cells of GOMTR1, *E. coli* HB101 (pRK2013), and *E. coli* JM109 (pSUPSN::Km) were collected separately by centrifugation and each cell suspension in LB  
10 medium was mixed in the ratio of 10:1:1, respectively. Then these cell suspensions were mixed at the same volume and the mixture was spread out on a 0.45 µm nitrocellulose membrane (PROTRAN, Schleicher & Schuell GmbH, Postfach 4, D-37582 Dassel, Germany) put on an agar medium, which was composed of 5.0% mannitol, 0.25% MgSO<sub>4</sub>·7H<sub>2</sub>O, 1.75% corn steep liquor, 5.0% baker's yeast, 0.5% urea, 0.5% CaCO<sub>3</sub>, and  
15 2.0% agar, for conjugal transfer of the suicide plasmid from the *E. coli* donor to GOMTR1 as recipient. After cultivation at 27°C for 1 day, the cells containing transconjugants were suspended and diluted appropriately with T broth, and spread out on the screening agar plates containing 100 µg/ml of rifampicin and 50 µg/ml of kanamycin. Finally, several transconjugants (Km<sup>r</sup>Rif<sup>r</sup>Tc<sup>s</sup>) which had the disrupted  
20 SNDH gene with Km cassette were obtained.

GOMTR1 and the disruptants, GOMTR1SN::Km, were grown on an agar plate containing 8.0% L-sorbose, 0.25% MgSO<sub>4</sub>·7H<sub>2</sub>O, 1.75% corn steep liquor, 5.0% baker's yeast, 0.5% urea, 0.5% CaCO<sub>3</sub>; and 2.0% agar at 27°C for 4 days. One loopful of the cells was inoculated into 50 ml of a seed culture medium (pH 6.0) containing 4% D-sorbitol,  
25 0.4% yeast extract, 0.05% glycerol, 0.25% MgSO<sub>4</sub>·7H<sub>2</sub>O, 1.75% corn steep liquor, 0.1% urea, and 1.5% CaCO<sub>3</sub> in a 500 ml Erlenmeyer flask, and cultivated at 30°C with 180 rpm for 1 day on a rotary shaker. The seed culture thus prepared was used for inoculating 50 ml of a main culture medium, which was composed of 12.0% L-sorbose, 2.0% urea, 0.05% glycerol, 0.25% MgSO<sub>4</sub>·7H<sub>2</sub>O, 3.0% corn steep liquor, 0.4% yeast extract, and  
30 1.5% CaCO<sub>3</sub> in a 500 ml Erlenmeyer flask. The cultivation was carried out at 30°C and 180 rpm for 4 days. The activity assay was performed as described in Example 4. The amount of 2-KGA produced as the enzyme activity was measured at a wavelength of 340nm by a HPLC system that was composed with a UV detector (TOSOH UV8000; Tosoh), a dualpump (TOSOH CCPE; Tosoh), an integrator (Shimadzu C-R6A; Shimadzu) and a column (YMC-Pack Pro C18, YMC). As shown in Table 2, the  
35

production efficiency for 2-KGA of the SNDH-gene disruptants was higher than that of the parent strain GOMTR1. The difference of the conversion rate per mol L-sorbose to 2-KGA was about 3%.

5 Table 2: 2-KGA production of *G. oxydans* strains having a disrupted SNDH-gene

Strain	2-KGA (g/L)	Residual L-sorbose (g/L)	*Molar yield (mol %)
GOMTR1SN::Km	96.7	15.3	99.2
GOMTR1	98.8	9.8	95.5

\*Molar yield: mol 2-KGA produced/mol L-sorbose consumed.

Example 6: Introduction of the plasmids carrying the SNDH gene into the SNDH-gene disruptant of *G. oxydans* DSM 4025

10 Several kinds of SNDH-expression plasmids using broad host range vector pVK100 were constructed as shown in Figure 5. Those plasmids have different insert DNAs at the *Hind* III site of pVK100 described as follows: pVSN117 has the insert DNA containing the incomplete SNDH gene encoding a polypeptide ending at Gly535 of SEQ ID NO: 5 (amino acid residue 566 of SEQ ID NO: 2), i.e., a C-terminal deleted SNDH gene, which  
 15 expresses only a 55 kDa protein. Plasmids pVSN106 and pVSN114, respectively, have the insert DNA containing the complete SNDH gene. Those plasmids were introduced into strain GOMTR1SN::Km by conjugal transfer method.

The transconjugants having the plasmids shown in Figure 5 were grown on an agar plate containing 10.0% L-sorbose, 0.25% MgSO<sub>4</sub>·7H<sub>2</sub>O, 1.75% corn steep liquor, 5.0%  
 20 baker's yeast, 0.5% urea, 0.5% CaCO<sub>3</sub>, and 2.0% agar at 27°C for 4 days. The enzyme reaction mixture consisted of 80 µg of cell free extract of the recombinant *Gluconobacter* strains, 25 mM potassium phosphate buffer (pH 7.0), 50 mM of L-sorbose, and 0.05 mM of PMS. The enzyme reaction was carried out at 30°C for 30 min with shaking at 1,000 rpm. The activity assay was performed according to Example 4. The result is  
 25 shown in Table 3.

Table 3: production of vitamin C with different SNDH constructs

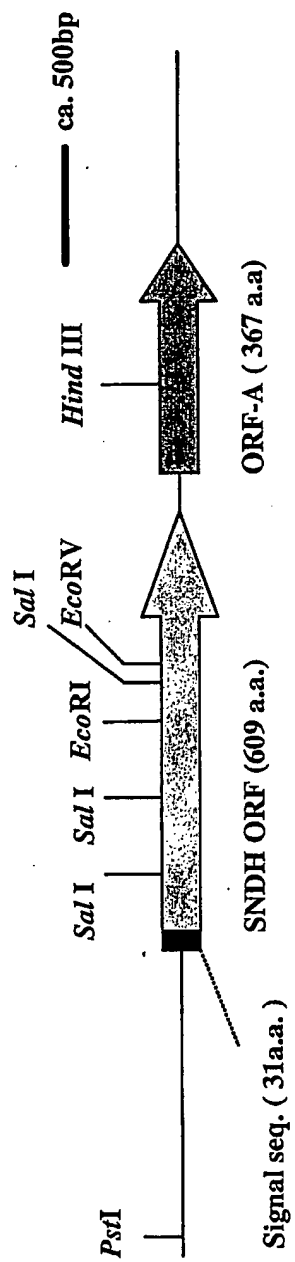
Host cell	Vector DNA	Vitamin C produced (mg/L)
GOMTR1SN::Km-2	pVK100	0.0
	pVSN117	473.2
	pVSN106	845.3
	pVSN114	860.2

Claims

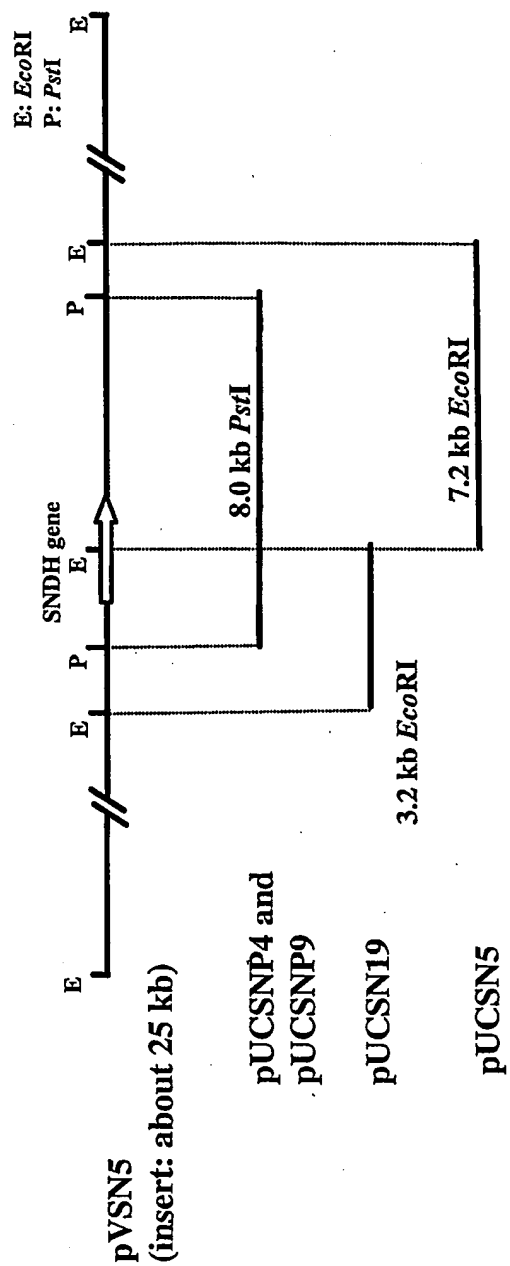
1. An isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide being at least 95% identical to the nucleotide sequence of SEQ ID NO: 1.
- 5 2. An isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide being at least 95% identical to the polynucleotide selected from the group consisting of (a) nucleotides 258-2084 of SEQ ID NO: 1, (b) nucleotides 351-2084 of SEQ ID NO: 1, (c) nucleotides 258-1955 of SEQ ID NO: 1, and (d) nucleotides 351-1955 of SEQ ID NO: 1.
- 10 3. An isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide selected from the group consisting of (a) a polynucleotide encoding the polypeptide having the amino acid sequence of SEQ ID NO: 2, (b) a polynucleotide encoding the polypeptide consisting of amino acids 32-609 of SEQ ID NO: 2, (c) a polynucleotide encoding the polypeptide consisting of amino acids 1-566 of  
15 SEQ ID NO: 2, and (d) a polynucleotide encoding the polypeptide consisting of amino acids 32-566 of SEQ ID NO: 2.
4. An isolated nucleic acid molecule encoding a polypeptide having aldehyde dehydrogenase activity, wherein said nucleic acid molecule hybridizes under standard conditions to the complementary strand of a nucleic acid molecule of any one of claims 1  
20 to 3.
5. An expression vector which comprises the nucleic acid molecule of any one of claims 1 to 4.
6. The expression vector of claim 5, wherein said vector is selected from a vector or a derivative thereof selected from the group consisting of pQE, pUC, pBluescript II,  
25 pACYC177, pACYC184, pVK100 and RSF1010.
7. A recombinant microorganism which is transformed with the expression vector of claim 5.
8. A recombinant microorganism which comprises the nucleic acid molecule of any one of claims 1 to 4 integrated into its chromosomal DNA.
- 30 9. The recombinant microorganism of claim 7 or 8, wherein said microorganism is selected from the group consisting of bacteria, yeast, and plant cells.

10. The recombinant microorganism of claim 9, wherein said microorganism is selected from the group consisting of *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Klebsiella*, *Acinetobacter*, and *Escherichia*.
11. The recombinant microorganism of claim 10, wherein said microorganism is  
5 *Gluconobacter oxydans* DSM 4025.
12. A process for the production of 2-keto-L-gulonic acid (2-KGA) and/or vitamin C from L-sorbose comprising (a) cultivating the recombinant microorganism of claim 7 or 8 in an appropriate culture medium, and (b) recovering and separating 2-KGA and/or vitamin C from said culture medium.
- 10 13. A process for the production of 2-KGA from L-sorbose comprising (a) cultivating a microorganism belonging to *Gluconobacter oxydans* DSM 4025 in an appropriate culture medium, wherein the gene encoding aldehyde dehydrogenase represented by SEQ ID NO: 2 is disrupted in said microorganism, and (b) recovering and separating 2-KGA from said culture medium.
- 15 14. A process for the production of aldehyde dehydrogenase comprising (a) cultivating a recombinant microorganism comprising a nucleic acid molecule of any one of claims 1 to 4 in an appropriate culture medium, and (b) recovering and separating said aldehyde dehydrogenase from said culture medium.

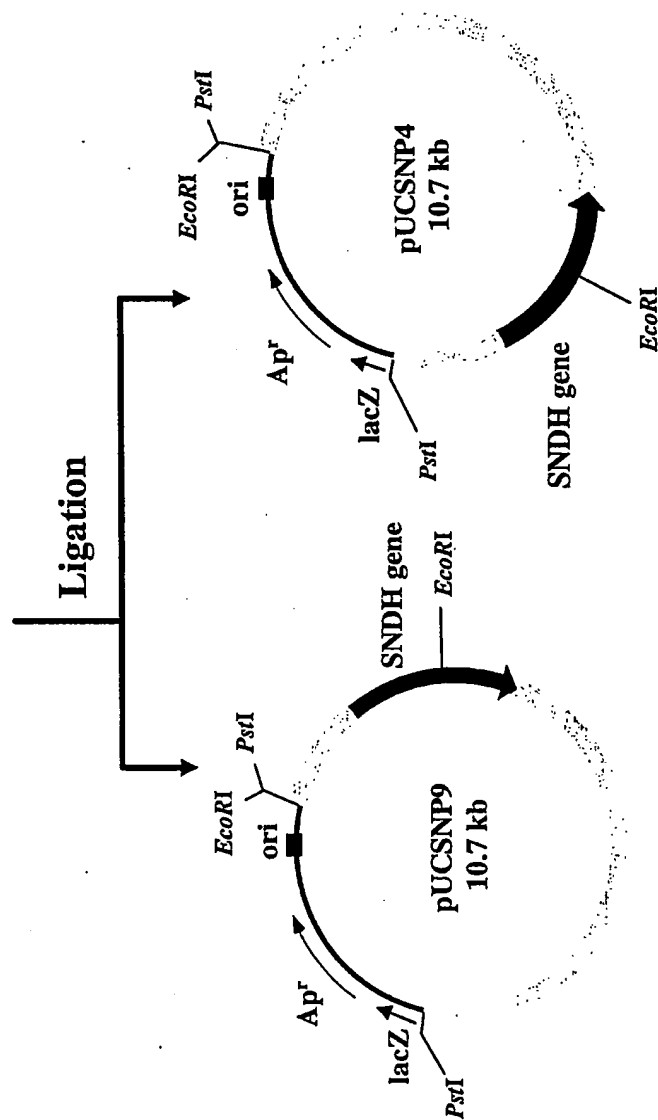
\*\*\*



**Fig. 1**

**Fig. 2**

- 8.0 kb *Pst*I fragment including intact SNDH gene from pVSN5
- pUC18 plasmid vector digested with *Pst*I



SNDH ORF-A NEW

**Fig. 3**

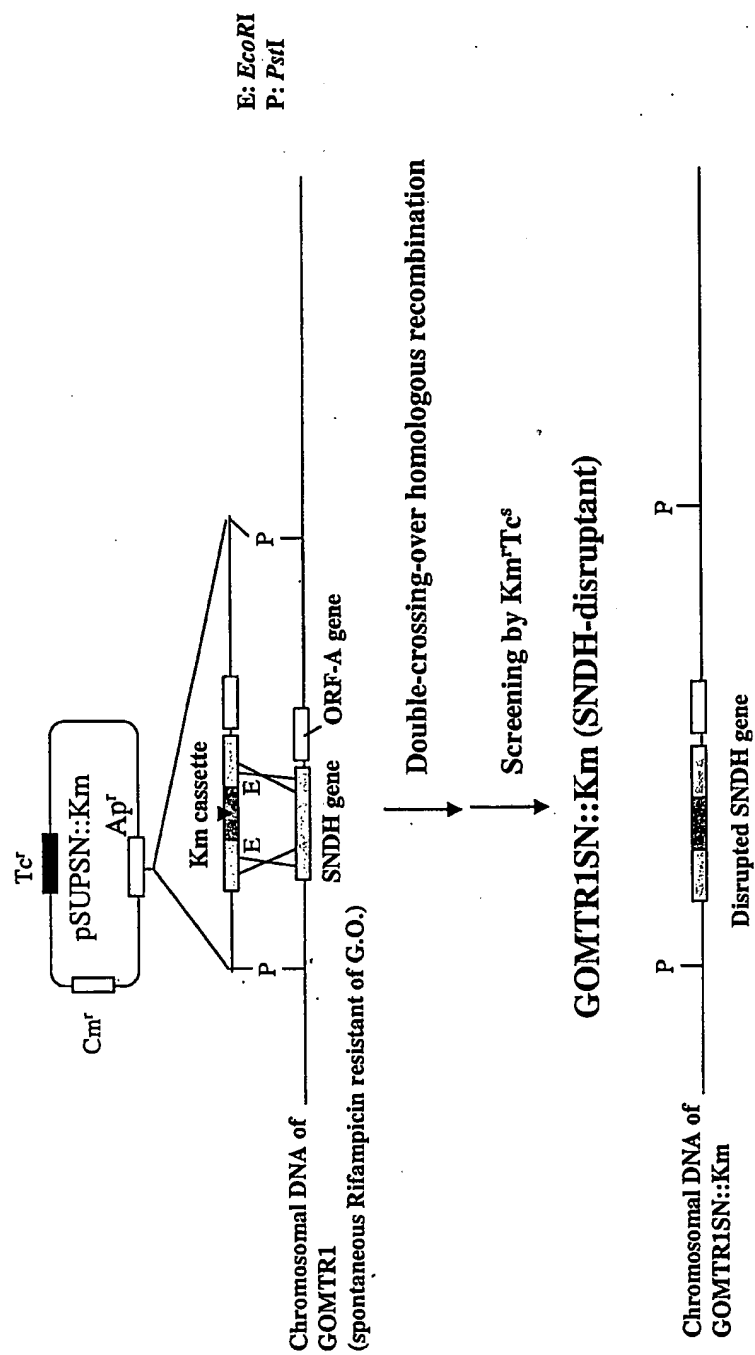


FIG. 4

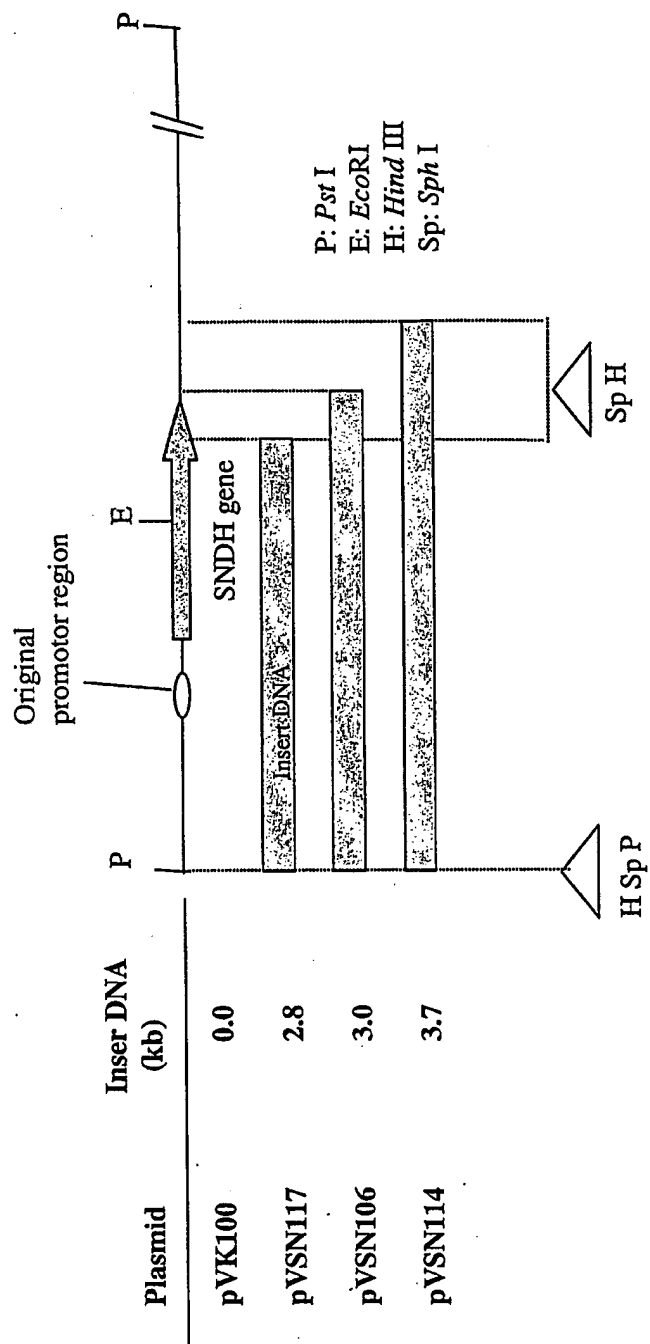


Fig. 5

- 1 -

Sequence Listings

<110> Roche Vitamins AG  
<120> A gene encoding aldehyde dehydrogenase and use thereof  
<130> 21424  
5 <140>  
<141>  
<160> 9  
<170> PatentIn Ver. 3.1

10 <210> 1  
<211> 3408  
<212> DNA  
<213> Gluconobacter oxydans  
<220>  
15 <221> CDS  
<222> (258)..(2087)  
<220>  
<221> CDS  
<222> (2214)..(3317)

20  
<400> 1  
GCGACTGGCA GCAGCGCAAC TATGACCACT ATGGCCTGCC GCCCTATTGG 50  
ATCTAACTGA TCCAGTAAGC CACCATCAGC CGGCCCTGC GGGGGCCGGC 100  
TTTTTGCCT AGACCCCGCC GAGGTGCTGT CGTAACCTAA GGTCACATCT 150  
25 TTAATTCCAC ATCCGCCCTT GTCAGTCTTG ACGTGACAAA TTGTCGCGGT 200  
CATGCTGCTG AATGCGGATG CCAGTCCCAG ATCCAAGCCC GACGCAAGGA 250  
GACGTAGATG TTACCCAAAT CATTGAAACA TAAGAATGGC GCCATGCGCC 300  
TTGTCGCAGC CTCGACCCTT GCGCTGATGA TCGGCGCGGG TGCCCATGCG 350  
CAGGTAAACC CGGTCGAAGT GCCGGTGGGC GCGAACGAGA CCTTTACCTC 400  
30 GCGCGTGCTG ACCACCGGCC TGTCGAACCC TTGGGAAATC ACCTGGGGCC 450  
CCGACAATAT GCTGTGGGTG ACCGAGCGAT CTTCCGGCGA AGTGACGCGC 500  
GTCGACCCCA ATACCGGCGA GCAGCAGGTC CTGCTGACCC TGACCGATTT 550  
CAGCGTCGAT GTGCAACACC AGGGCCTACT TGGCCTCGCG CTGCATCCTG 600  
AGTTTATGCA AGAGAGCGGC AACGACTACG TCTATATCGT CTACACTTAT 650  
35 AACACCGGCA CCGAAGAAGC GCCCGATCCG CATCAAAGC TGGTGCGTTA 700  
TGCTTATGAC GCTGCCGCGC AGCAGCTGGT CGATCCGGTT GATCTGGTGC 750  
CAGGCATTCC CGCAGGCAAC GACCAACAATG GCGGTCGCAT CAAATTCGCC 800  
CCCGATGGCC AACACATCTT TTACACGCTG GGCGAGCAAG GCGCGAAGTT 850  
TGGCGGTAAC TTCCGCCGTC CGAACCACGC GCAACTGCTG CCGACGCAAG 900  
40 AGCAGGTCGA CGCGGGCGAT TGGGTCGCCT ATTCGGGCAA GATCCTGCGC 950

- 2 -

GTGAACCTTG ACGGCACGAT CCCCGAAGAC AACCCCGAGA TCGAGGGCGT 1000  
 GCGTAGCCAT ATCTTTACCT ATGGCCACCG TAACCCGCAG GGCATCACCT 1050  
 TTGGCCCCGA CGGCACCATT TATGCCACCG AACACGGCCC CGATACGGAT 1100  
 GACGAGCTGA ACATCATCGC CGGCGGTGGC AACTATGGGT GGCCGAATGT 1150  
 5 GGCCGGCTAT CGCGATGGCA AATCCTATGT CTACGCTGAT TGGAGCCAAG 1200  
 CGCCCGCTGA CCAGCGTTAC ACCGGTCGCG CCGGTATCCC CGACACCGTG 1250  
 CCGCAATTCC CCGAGCTGGA ATTCGCGCCC GAGATGGTCG ATCCGCTGAC 1300  
 AACCTATTGG ACGGTGGATA ATGATTACGA TTTCACCGCC AATTGCGGCT 1350  
 GGATCTGTAA TCCGACGATC GCGCCTTCGT CTGCCTATTA CTATGCGGCG 1400  
 10 GGCGAGAGCG GTATCGCGGC TTGGGATAAT TCGATCCTGA TCCCGACGCT 1450  
 GAAACATGGC GGCATCTATG TGCAGCACCT CAGCGATGAT GGCCAATCTG 1500  
 TCGACGGCCT GCCCGAGCTG TGGTTCAGCA CCCAGAACCG CTATCGCGAT 1550  
 ATCGAGATCA GCCCCGATAA CCATGTTTTT GTGGCGACCG ACAACTTTGG 1600  
 CACCTCGGCG CAGAAATATG GCGAGACCGG CTTTACCAAC GTGCTGCATA 1650  
 15 ACCCCGGCGC GATCCTTGTC TTTAGCTATG TCGGCGAGGA TGCTGCGGGT 1700  
 CAGACCGGAA TGATGACCGC GCCCGCACCG CAGACGCAAT ACACGCAAGT 1750  
 GCCCGCCGAG GGTGCAGGCG CGGGCGCGAC TGAGGTTGCG GATGTCGATT 1800  
 ACGACACGCT GTTCACCGAA GGCCAGACCC TTTATGGCAG CGCATGTGCC 1850  
 GCGTGCCATG GTGCCGCTGG CCAAGGTGCG CAGGGCCCCA CCTTTGTGGG 1900  
 20 CGTGCCGGAT GTGACGGGTG ACAAGGACTA CCTTGCCCGC ACCATCATCC 1950  
 ACGGTTTTGG CTATATGCCG TCGTTTGCGA CTCGGCTGGA TGACGAGGAG 2000  
 GTTGCCGCCA TCGCGACCTT TATCCGCAAC AGCTGGGGCA ATGACGAAGG 2050  
 CATCCTGACC CCGCCGAGG CCGCTGCCAC CCGCTGAATG CTGTAAAAAC 2100  
 CACCCTCGCC TGCACATCAG GCGGGGGTAT TTCATTTATT TTCACATCTG 2150  
 25 CCTTTGACAT GTGCCGCTAT CACGGTTAAT GCGGCCCTTC GGCTGTTCTG 2200  
 GGTCTAAGCG GGTGTGTTGC CCGATAAGAG AGACGGTTCA GTCCCTCCCG 2250  
 CCCTATTTAG GGCCATTTA GGAGAATAG TTTTGACTCA TCAAAATATC 2300  
 GCCGCGCCTC TGGCCGCGGC CCTTTCGCAA CGTGGATATG AAACGCTGAC 2350  
 CGCCGTGCAG CAAGCTGTGC TTGCGCCCGA GGCTGATGGC CGCGACCTGC 2400  
 30 TGGTGTGCGC ACAGACCGGT TCGGTAAGA CCGTGCCCTT TGGTATCGCA 2450  
 GTCGCGCCCG ACCTTTTGGG CGACGACAAT ATCCTGCCGC TGAACACGCC 2500  
 GCCTGTTGCG CTGTTTCATG CCCCCACGCG CGAGCTTGCG CTGCAAGTTG 2550  
 CTCAGGAAGT GACCTGGCTT TACGCCAATG CAGGTGCCCA GATCGCGACC 2600  
 TGCGTCGGCG GTATGGATTA CCGCACCGAG CGCCGCGCCC TTGCACGTCT 2650  
 35 GCCGCAATC GTTGTGCGCA CGCCCGGCCG TCTGCGCGAC CATATCGACC 2700  
 GTGGCGGCCT TGACCTGTCC GAATTGCGCG TGACCGTGCT GGACGAAGCG 2750  
 GATGAGATGC TCGACCTCGG CTTCCGCGAT GATCTGCAAT ATATCTTGCA 2800  
 AGCCGCGCCC GAAGATCGCC GCACGCTGAT GTTCTCGGCC ACCGTGCCGC 2850  
 GCGAGATTGA AAAACTGGCC CGCGACTTCC AAAATGACGC CCTGCGTCTG 2900  
 40 GAAACCCGTG GCGAGGCCAA GCAGCACAAC GACATCAGCT ACCAAGCTTT 2950  
 GTCGGTCACC ATGCGCGATC GCGAAAACGC CATTTTCAAC ATGCTGCGTT 3000

- 3 -

TTTATGAATC GCGCACGGCG ATCATCTTCT GCAAGACCCG CGCCAATGTG 3050  
 AATGATCTGC TGTCGCGGAT GAGCGGTCGT GGCTTCCGCG TGGTGGCCCT 3100  
 GTCGGGCGAG CTGTCGCAAC AGGAACGCAC CAACGCGCTG CAAGCGCTGC 3150  
 GTGATGGCCG CGCCAACGTT TGTATCGCGA CCGACGTCGC GGCGCGCGGC 3200  
 5 ATTGACTTGC CGGGCCTCGA GCTGGTGATC CACTACGATC TGCCGACCAA 3250  
 TGCCGAAACC CTGCTGCACC GCTCGGGCCG TACCGGCCGC CGGGTGCCAA 3300  
 GGGCGTCTCG GCGCTGATCG TCACCCCGG CGATTTCAAA AAAGCGCAGC 3350  
 GTTTGCTGAG CTTTGCCAAA GTGACCGCGG AATGGGGCAA GGCGCCTTCG 3400  
 GCCGAAGA 3408

10

<210> 2  
 <211> 609  
 <212> PRT  
 15 <213> Gluconobacter oxydans  
 <220>  
 <221> SIGNAL  
 <222> (1) .. (31)  
 <220>  
 20 <221> CHAIN  
 <222> (32) .. (609)

<400> 2  
 Met Leu Pro Lys Ser Leu Lys His Lys Asn Gly Ala Met Arg Leu 15  
 25 Val Ala Ala Ser Thr Leu Ala Leu Met Ile Gly Ala Gly Ala His 30  
 Ala Gln Val Asn Pro Val Glu Val Pro Val Gly Ala Asn Glu Thr 45  
 30 Phe Thr Ser Arg Val Leu Thr Thr Gly Leu Ser Asn Pro Trp Glu 60  
 Ile Thr Trp Gly Pro Asp Asn Met Leu Trp Val Thr Glu Arg Ser 75  
 Ser Gly Glu Val Thr Arg Val Asp Pro Asn Thr Gly Glu Gln Gln 90  
 35 Val Leu Leu Thr Leu Thr Asp Phe Ser Val Asp Val Gln His Gln 105  
 Gly Leu Leu Gly Leu Ala Leu His Pro Glu Phe Met Gln Glu Ser 120  
 40 Gly Asn Asp Tyr Val Tyr Ile Val Tyr Thr Tyr Asn Thr Gly Thr 135  
 Glu Glu Ala Pro Asp Pro His Gln Lys Leu Val Arg Tyr Ala Tyr 150

- 4 -

Asp Ala Ala Ala Gln Gln Leu Val Asp Pro Val Asp Leu Val Ala 165  
Gly Ile Pro Ala Gly Asn Asp His Asn Gly Gly Arg Ile Lys Phe 180  
5 Ala Pro Asp Gly Gln His Ile Phe Tyr Thr Leu Gly Glu Gln Gly 195  
Ala Asn Phe Gly Gly Asn Phe Arg Arg Pro Asn His Ala Gln Leu 210  
10 Leu Pro Thr Gln Glu Gln Val Asp Ala Gly Asp Trp Val Ala Tyr 225  
Ser Gly Lys Ile Leu Arg Val Asn Leu Asp Gly Thr Ile Pro Glu 240  
Asp Asn Pro Glu Ile Glu Gly Val Arg Ser His Ile Phe Thr Tyr 255  
15 Gly His Arg Asn Pro Gln Gly Ile Thr Phe Gly Pro Asp Gly Thr 270  
Ile Tyr Ala Thr Glu His Gly Pro Asp Thr Asp Asp Glu Leu Asn 285  
20 Ile Ile Ala Gly Gly Gly Asn Tyr Gly Trp Pro Asn Val Ala Gly 300  
Tyr Arg Asp Gly Lys Ser Tyr Val Tyr Ala Asp Trp Ser Gln Ala 315  
Pro Ala Asp Gln Arg Tyr Thr Gly Arg Ala Gly Ile Pro Asp Thr 330  
25 Val Pro Gln Phe Pro Glu Leu Glu Phe Ala Pro Glu Met Val Asp 345  
Pro Leu Thr Thr Tyr Trp Thr Val Asp Asn Asp Tyr Asp Phe Thr 360  
30 Ala Asn Cys Gly Trp Ile Cys Asn Pro Thr Ile Ala Pro Ser Ser 375  
Ala Tyr Tyr Tyr Ala Ala Gly Glu Ser Gly Ile Ala Ala Trp Asp 390  
Asn Ser Ile Leu Ile Pro Thr Leu Lys His Gly Gly Ile Tyr Val 405  
35 Gln His Leu Ser Asp Asp Gly Gln Ser Val Asp Gly Leu Pro Glu 420  
Leu Trp Phe Ser Thr Gln Asn Arg Tyr Arg Asp Ile Glu Ile Ser 435  
40 Pro Asp Asn His Val Phe Val Ala Thr Asp Asn Phe Gly Thr Ser 450  
Ala Gln Lys Tyr Gly Glu Thr Gly Phe Thr Asn Val Leu His Asn 465

- 5 -

Pro Gly Ala Ile Leu Val Phe Ser Tyr Val Gly Glu Asp Ala Ala 480  
Gly Gln Thr Gly Met Met Thr Ala Pro Ala Pro Gln Thr Gln Tyr 495  
5 Thr Gln Val Pro Ala Glu Gly Ala Gly Ala Gly Ala Thr Glu Val 510  
Ala Asp Val Asp Tyr Asp Thr Leu Phe Thr Glu Gly Gln Thr Leu 525  
10 Tyr Gly Ser Ala Cys Ala Ala Cys His Gly Ala Ala Gly Gln Gly 540  
Ala Gln Gly Pro Thr Phe Val Gly Val Pro Asp Val Thr Gly Asp 555  
Lys Asp Tyr Leu Ala Arg Thr Ile Ile His Gly Phe Gly Tyr Met 570  
15 Pro Ser Phe Ala Thr Arg Leu Asp Asp Glu Glu Val Ala Ala Ile 585  
Ala Thr Phe Ile Arg Asn Ser Trp Gly Asn Asp Glu Gly Ile Leu 600  
20 Thr Pro Ala Glu Ala Ala Ala Thr Arg 609

<210> 3  
<211> 14  
25 <212> PRT  
<213> Gluconobacter oxydans

<400> 3  
Gln [Xaa/Gly] Asn [Pro/Lys] Val Glu Val Pro Val Gly Ala Asn Glu Thr 14  
30

<210> 4  
<211> 31  
<212> PRT  
35 <213> Gluconobacter oxydans

<220>  
<221> SIGNAL  
<222> (1) .. (31)  
40  
<400> 4

- 6 -

Met Leu Pro Lys Ser Leu Lys His Lys Asn Gly Ala Met Arg Leu 15

Val Ala Ala Ser Thr Leu Ala Leu Met Ile Gly Ala Gly Ala His 30

5 Ala 31

<210> 5

<211> 578

10 <212> PRT

<213> Gluconobacter oxydans

<220>

<221> CHAIN

15 <222> (1) .. (578)

<400> 5

Gln Val Asn Pro Val Glu Val Pro Val Gly Ala Asn Glu Thr Phe 15

20 Thr Ser Arg Val Leu Thr Thr Gly Leu Ser Asn Pro Trp Glu Ile 30

Thr Trp Gly Pro Asp Asn Met Leu Trp Val Thr Glu Arg Ser Ser 45

Gly Glu Val Thr Arg Val Asp Pro Asn Thr Gly Glu Gln Gln Val 60

25 Leu Leu Thr Leu Thr Asp Phe Ser Val Asp Val Gln His Gln Gly 75

Leu Leu Gly Leu Ala Leu His Pro Glu Phe Met Gln Glu Ser Gly 90

30 Asn Asp Tyr Val Tyr Ile Val Tyr Thr Tyr Asn Thr Gly Thr Glu 105

Glu Ala Pro Asp Pro His Gln Lys Leu Val Arg Tyr Ala Tyr Asp 120

Ala Ala Ala Gln Gln Leu Val Asp Pro Val Asp Leu Val Ala Gly 135

35 Ile Pro Ala Gly Asn Asp His Asn Gly Gly Arg Ile Lys Phe Ala 150

Pro Asp Gly Gln His Ile Phe Tyr Thr Leu Gly Glu Gln Gly Ala 165

40 Asn Phe Gly Gly Asn Phe Arg Arg Pro Asn His Ala Gln Leu Leu 180

- 7 -

Pro Thr Gln Glu Gln Val Asp Ala Gly Asp Trp Val Ala Tyr Ser 195

Gly Lys Ile Leu Arg Val Asn Leu Asp Gly Thr Ile Pro Glu Asp 210

5 Asn Pro Glu Ile Glu Gly Val Arg Ser His Ile Phe Thr Tyr Gly 225

His Arg Asn Pro Gln Gly Ile Thr Phe Gly Pro Asp Gly Thr Ile 240

Tyr Ala Thr Glu His Gly Pro Asp Thr Asp Asp Glu Leu Asn Ile 255

10 Ile Ala Gly Gly Gly Asn Tyr Gly Trp Pro Asn Val Ala Gly Tyr 270

Arg Asp Gly Lys Ser Tyr Val Tyr Ala Asp Trp Ser Gln Ala Pro 285

15 Ala Asp Gln Arg Tyr Thr Gly Arg Ala Gly Ile Pro Asp Thr Val 300

Pro Gln Phe Pro Glu Leu Glu Phe Ala Pro Glu Met Val Asp Pro 315

Leu Thr Thr Tyr Trp Thr Val Asp Asn Asp Tyr Asp Phe Thr Ala 330

20 Asn Cys Gly Trp Ile Cys Asn Pro Thr Ile Ala Pro Ser Ser Ala 345

Tyr Tyr Tyr Ala Ala Gly Glu Ser Gly Ile Ala Ala Trp Asp Asn 360

25 Ser Ile Leu Ile Pro Thr Leu Lys His Gly Gly Ile Tyr Val Gln 375

His Leu Ser Asp Asp Gly Gln Ser Val Asp Gly Leu Pro Glu Leu 390

Trp Phe Ser Thr Gln Asn Arg Tyr Arg Asp Ile Glu Ile Ser Pro 405

30 Asp Asn His Val Phe Val Ala Thr Asp Asn Phe Gly Thr Ser Ala 420

Gln Lys Tyr Gly Glu Thr Gly Phe Thr Asn Val Leu His Asn Pro 435

35 Gly Ala Ile Leu Val Phe Ser Tyr Val Gly Glu Asp Ala Ala Gly 450

Gln Thr Gly Met Met Thr Ala Pro Ala Pro Gln Thr Gln Tyr Thr 465

Gln Val Pro Ala Glu Gly Ala Gly Ala Gly Ala Thr Glu Val Ala 480

40 Asp Val Asp Tyr Asp Thr Leu Phe Thr Glu Gly Gln Thr Leu Tyr 495

- 8 -

Gly Ser Ala Cys Ala Ala Cys His Gly Ala Ala Gly Gln Gly Ala 510

Gln Gly Pro Thr Phe Val Gly Val Pro Asp Val Thr Gly Asp Lys 525

5 Asp Tyr Leu Ala Arg Thr Ile Ile His Gly Phe Gly Tyr Met Pro 540

Ser Phe Ala Thr Arg Leu Asp Asp Glu Glu Val Ala Ala Ile Ala 555

10 Thr Phe Ile Arg Asn Ser Trp Gly Asn Asp Glu Gly Ile Leu Thr 570

Pro Ala Glu Ala Ala Ala Thr Arg 578

15 <210> 6  
<211> 17  
<212> DNA  
<213> Artificial Sequence

20 <220>  
<223> an artificially synthesized primer sequence

<400> 6  
carggyaacc csgtbga 17

25 <210> 7  
<211> 17  
<212> DNA  
<213> Artificial Sequence

30 <220>  
<223> an artificially synthesized primer sequence

<220>

35 <221> misc\_feature  
<222> 9  
<223> n is a or g or c or t

<400> 7  
40 gtytcgttng crccvac 17

- 9 -

<210> 8  
<211> 15  
<212> DNA  
5 <213> Artificial Sequence

<220>  
<223> an artificially synthesized primer sequence

10 <400> 8  
cagggtacc cggtc 15

<210> 9  
15 <211> 15  
<212> DNA  
<213> Artificial Sequence

<220>  
20 <223> an artificially synthesized primer sequence


<400> 9  
gactcgtttg cgccc 15

**BUDAPEST TREATY ON THE INTERNATIONAL  
RECOGNITION OF THE DEPOSIT OF MICROORGANISMS  
FOR THE PURPOSES OF PATENT PROCEDURE**

**INTERNATIONAL FORM**

The Oriental Scientific  
Instruments, Import and Export  
Corp. for Inst. of Microbiology  
Academia Sinica  
52 San-Li-He Rd.  
Beijing, P.R. China

**VIABILITY STATEMENT**  
issued pursuant to Rule 10.2 by the  
INTERNATIONAL DEPOSITARY AUTHORITY  
identified at the bottom of this page

<b>I. DEPOSITOR</b>		<b>II. IDENTIFICATION OF THE MICROORGANISM</b>			
Name:	The Oriental Scientific Instruments, Import and Export	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY:			
Address:	Corp. for Inst. of Microbiology Academia Sinica 52 San-Li-He Rd. Beijing, P.R. China	DSM 4025			
		Date of the deposit or the transfer <sup>1</sup> :			
		1987-03-17			
<b>III. VIABILITY STATEMENT</b>					
The viability of the microorganism identified under II above was tested on 2003-08-13 <sup>2</sup> . On that date, the said microorganism was					
<input checked="" type="checkbox"/> viable					
<input type="checkbox"/> no longer viable					
<b>IV. CONDITIONS UNDER WHICH THE VIABILITY TEST HAS BEEN PERFORMED<sup>4</sup></b>					
<b>V. INTERNATIONAL DEPOSITARY AUTHORITY</b>					
Name:	DSMZ-DEUTSCHE SAMMLUNG VON MIKROORGANISMEN UND ZELLKULTUREN GmbH	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s):			
Address:	Mascheroder Weg 1b D-38124 Braunschweig				
		Date: 2003-08-15			

<sup>1</sup> Indicate the date of original deposit or, where a new deposit or a transfer has been made, the most recent relevant date (date of the new deposit or date of the transfer).

<sup>2</sup> In the cases referred to in Rule 10.2(a) (ii) and (iii), refer to the most recent viability test.

<sup>3</sup> Mark with a cross the applicable box.

<sup>4</sup> Fill in if the information has been requested and if the results of the test were negative.

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
8 April 2004 (08.04.2004)

PCT

(10) International Publication Number  
**WO 2004/029235 A3**

- (51) International Patent Classification<sup>7</sup>: C12N 9/04, 15/52, 5/10, C12P 7/60
- (21) International Application Number: PCT/EP2003/010498
- (22) International Filing Date: 22 September 2003 (22.09.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 02021650.3 27 September 2002 (27.09.2002) EP
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant (*for all designated States except US*): DSM IP ASSETS B.V. [NL/NL]; Het Overloon 1, NL-6411 TE Heerlen (NL).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): HOSHINO, Tatsuo [JP/JP]; 2-18-14 Fueta, Kamakura-shi, Kanagawa-ken 248-0027 (JP). MIYAZAKI, Taro [JP/JP]; Kameino 2-14-8, Fujisawa-shi, Kanagawa-ken 252-0813 (JP). SUGISAWA, Teruhide [JP/CH]; Inzlingerstrasse 80, CH-4125 Riehen (CH).
- (74) Agent: SCHWANDER, Kuno, Josef; Roche Vitamins Ltd., Patent Department (VMD), Wurmisweg 576, CH-4303 Kaiseraugst (CH).
- Published:**
- with international search report
  - before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
  - with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description
- (88) Date of publication of the international search report: 8 July 2004
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: ALDEHYDE DEHYDROGENASE GENE

(57) Abstract: The present invention relates to a DNA which encodes aldehyde dehydrogenase (SNDH), an expression vector containing the DNA and recombinant organisms containing said DNA. Furthermore, the present invention concerns a process for producing recombinant aldehyde dehydrogenase protein and a process for producing L-ascorbic acid (vitamin C) and/or 2-keto-L-gulonic acid (2-KGA) from L-sorbose by using the recombinant aldehyde dehydrogenase protein or recombinant organisms containing the expression vector. Also provided is a process for the production of 2-KGA with a microorganism in which the gene encoding said aldehyde dehydrogenase is disrupted.

WO 2004/029235 A3

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/EP 03/10498

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N9/04 C12N15/52 C12N5/10 C12P7/60

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

PAJ, EPO-Internal, EMBL, BIOSIS, WPI Data, SEQUENCE SEARCH, EMBASE

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SAITO Y ET AL: "Direct fermentation of 2-Keto-L-gulonic acid in recombinant <i>Gluconobacter oxydans</i> " BIOTECHNOLOGY AND BIOENGINEERING. INCLUDING: SYMPOSIUM BIOTECHNOLOGY IN ENERGY PRODUCTION AND CONSERVATION, JOHN WILEY & SONS. NEW YORK, US, vol. 58, no. 2-3, 20 April 1998 (1998-04-20), pages 309-315, XP002204789 ISSN: 0006-3592 figure 4  --- -/--	1-14



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*Z\* document member of the same patent family

Date of the actual completion of the international search

31 March 2004

Date of mailing of the international search report

16/04/2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Schmitz, T

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 03/10498

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SAITO Y ET AL: "CLONING OF GENES CODING FOR L-SORBOSE AND L-SORBOSONE DEHYDROGENASES FROM GLUCONOBACTER OXYDANS AND MICROBIAL PRODUCTION OF 2-KETO-L-GULONATE, A PRECURSOR OF L-ASCORBIC ACID, IN A RECOMBINANT G. OXYDANS STRAIN" APPLIED AND ENVIRONMENTAL MICROBIOLOGY, WASHINGTON,DC, US, vol. 63, no. 2, 1997, pages 454-460, XP000886144 ISSN: 0099-2240 the whole document	1-14
X	HOSHINO T ET AL: "ISOLATION AND CHARACTERIZATION OF NADP-DEPENDENT L SORBOSONE DEHYDROGENASE FROM GLUCONOBACTER-MELANOGENUS UV10" AGRICULTURAL AND BIOLOGICAL CHEMISTRY, vol. 55, no. 3, 1991, pages 665-670, XP002275642 ISSN: 0002-1369 the whole document	1-14
X	US 5 437 989 A (ASAKURA AKIRA ET AL) 1 August 1995 (1995-08-01) claim 1; examples 1-3	1-14
X	ASAKURA A ET AL: "ISOLATION AND CHARACTERIZATION OF A NEW QUINOPROTEIN DEHYDROGENASE, L-SORBOSE/L-SORBOSONE DEHYDROGENASE" BIOSCIENCE BIOTECHNOLOGY BIOCHEMISTRY, JAPAN SOC. FOR BIOSCIENCE, BIOTECHNOLOGY AND AGROCHEM. TOKYO, JP, vol. 63, no. 1, January 1999 (1999-01), pages 46-53, XP001085040 ISSN: 0916-8451 the whole document	1-14
X	EP 0 373 181 A (HOFFMANN LA ROCHE) 20 June 1990 (1990-06-20) the whole document	1-14
X	EP 0 790 301 A (HOFFMANN LA ROCHE) 20 August 1997 (1997-08-20) the whole document	1-14
A	WO 02 34919 A (HOFFMANN LA ROCHE ; ROCHE DIAGNOSTICS GMBH (DE)) 2 May 2002 (2002-05-02) SEQ ID NOs 24, 25	

-/--

# INTERNATIONAL SEARCH REPORT

International Application No.

PCT/EP 03/10498

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>JP 2002 125689 A (HAYADE KOJI)  8 May 2002 (2002-05-08)  SEQ ID NO:1</p>	
A	<p>-----  HANCOCK R D ET AL: "Biotechnological  approaches for l-ascorbic acid production"  TRENDS IN BIOTECHNOLOGY, ELSEVIER  PUBLICATIONS, CAMBRIDGE, GB,  vol. 20, no. 7, 1 July 2002 (2002-07-01),  pages 299-305, XP004361398  ISSN: 0167-7799  the whole document  -----</p>	

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 03/10498

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 5437989	A	01-08-1995	AT 224949 T	15-10-2002
			DE 69332331 D1	31-10-2002
			DE 69332331 T2	22-05-2003
			DK 606621 T3	06-01-2003
			EP 0606621 A2	20-07-1994
			ES 2181681 T3	01-03-2003
			JP 7000182 A	06-01-1995
			US 5932463 A	03-08-1999
			US 5916785 A	29-06-1999
EP 0373181	A	20-06-1990	DE 68924761 D1	14-12-1995
			DE 68924761 T2	04-04-1996
			DK 454689 A	14-09-1989
			EP 0373181 A1	20-06-1990
			US 5352599 A	04-10-1994
			AT 130030 T	15-11-1995
			WO 8906688 A2	27-07-1989
			JP 2799380 B2	17-09-1998
			JP 3500844 T	28-02-1991
EP 0790301	A	20-08-1997	EP 0790301 A2	20-08-1997
			AT 231183 T	15-02-2003
			CN 1171441 A ,B	28-01-1998
			DE 69718360 D1	20-02-2003
			DE 69718360 T2	16-10-2003
			DK 790301 T3	05-05-2003
			ES 2188808 T3	01-07-2003
			ID 15955 A	21-08-1997
			JP 9224660 A	02-09-1997
			US 5776742 A	07-07-1998
WO 0234919	A	02-05-2002	AU 1596602 A	06-05-2002
			BR 0114962 A	28-10-2003
			CA 2427029 A1	02-05-2002
			WO 0234919 A1	02-05-2002
			EP 1332216 A1	06-08-2003
			US 2003104595 A1	05-06-2003
			US 2004005683 A1	08-01-2004
JP 2002125689	A	08-05-2002	NONE	